Lab #3: More Relational Data Processing

Everything Data
CompSci 290.01 Spring 2014
Announcements (Tue. Jan. 28)

• Dan will be our only UTA for now
  – Office hours 9-11pm on Mondays in LINK
• We are still working on the Amazon VM…
  stay tuned
Seat assignment

Front of D106

Course staff

A

B

C

D

E

F

G

H

I

J

K

L

M

N

Back of D106
Format of this lab

• Class discussion of HW #3, Parts 1 & 2 + a few more SQL tricks (10 min.)
• Lab #3, Part 1 + team discussion of HW #3, Part 3 (15 minutes)
  – Share your claims and counterarguments
• Lab #3, Part 2: team challenge! (40 min.)
  – Win prizes and extra credit!
• Summarize lessons learned (10 min.)
HW #3, Part 2:

**Price vs. Pelosi/Cantor**

- There is a Republican Representative whose last name is also Price
  - The filtering condition in the query shown in Lecture #3 wasn’t precise enough
- Once we make the filtering condition precise, the result becomes expected:
  - David Price vs. Pelosi: 87%
  - David Price vs. Cantor: 31%
HW #3, Part 1 (A-D)

• Find Kennedy’s: filter
• Find Senate Kennedy’s: join
• Count reps by states: group-by
• Which Dems voted to repeal Obama Care?
  – In WHERE, use a subquery returning the vote_id of interest
  – In subquery, use LIKE or ~ for string pattern matching
HW #3, Part 1 (E)

Breakdown of votes by party

- Key: given vote_id and person_id, find the person’s party at the time of the vote

... (SELECT DISTINCT party
   FROM person_roles
   WHERE person_id = person_votes.person_id
   AND start_date <= votes.date AND votes.date <= end_date) ...

- Govtrack data isn’t clean: multiple roles may be active at the same time!

- **DISTINCT** option for SELECT eliminates duplicates
  
  - **DISTINCT** is also available for aggregation functions
    
    - E.g., COUNT(DISTINCT party) = # distinct parties in a group
Some tips for writing SQL

**Top-down**
- First write the “outer” query
- Then fill in the blanks with subqueries

**Bottom-up**
- For each output row, think about what evidence (rows from input tables) you need to prove it
  - That will give you the tables to join in FROM

*Think logically, on whole sets*
Queries too slow?

• **Indexing** may help
  – Index on a column speeds up looking up rows with a particular value (or within a range of values) for that column
  – E.g., `CREATE INDEX ON persons_roles(person_id);`

• *Note: if a column is PRIMARY KEY or UNIQUE, the database system will automatically create an index on it*
Team work (15 min.)

- Lab #3, Part 1
- Discuss HW #3, Part 3
  - Share your claims and counterarguments
- At the end of this period, I will give the signal to start the challenge
Team claim-checking challenge

• 4 claims, chosen from your submissions
• For each claim
  – Is it technically “correct”? 
  – Is it misleading? Why?
  – Show us your supporting queries

• Each claim successfully checked = extra credit worth 5% of a homework
• The first team to finish checking all claims gets a prize
Challenge!

• On my mark, refresh Lab #3 page for the list of claims

• Once your team finishes checking a claim, raise your hands and yell “CHECK ⟨claim #⟩!”
  – One of the course staff will come and verify your answer

• If we chose the claim from you, your team gets a freebie; no need to get it verified
A nation gone liberal?

The total length of terms served by Republicans has gone up by 135% from the 19th century (between 1800 and 1900) to the 20th century (between 1900 and 2000). But over the same period, the total length of terms served by Democrats has increased by a whopping 236%!

Based on submission by Eric Wu

• True, but historically Democrats were not always associated with liberalism
Democrats: a party of “NO”?

In 2013, Democratic members of the Congress casted more Nay/No votes than the Republicans. Based on submission by Prithvi Prabahar

• It depends on who controls the agenda
  – Try break down the votes further by chamber, and you will see that Senate Democrats are far more “positive”
Women on Obama Care

Females in the House are strong Obama Care supporters. Less than 1 in 4 females voted to repeal Obama Care. In contrast, more than 60% of the males voted to repeal it.

By Yi Hong Poo

• Once you break down the votes by party, you will see that women are basically voting on party lines as well
Chiesa the Republican

Jeffrey Chiesa, a Republican senator from New Jersey, voted fewer than 60 times with his party majority in 2013. That’s abysmal in comparison with Mitch McConnell, the Republican senate minority leader, who voted 262 times with the party majority.

By Morgan Krey

• He only served for 6 months (to replace the seat vacated by Lautenberg)
Lessons learned

• Knowing SQL doesn’t mean you know how to query a dataset

• Behold the power of simple but flexible data model + “declarative” queries
  – Data = collection of tables
  – Queries = compositions of filter, group-by aggregation, join, sort, subqueries, …

• *Physical data independence*
When to use a SQL database

• Pros and cons?
• Despite its age, SQL is still your best bet for querying your structured data
Finally

• Remember to submit team.txt by midnight
• Sample solutions to Lab #3 will be posted by tomorrow morning
  – Take some time on your own to go through the sample solutions